Reaching the Holevo Capacity via von Neumann measurement, and its use

Farkas Lóránt

February 1, 2008

1 Introduction

The method of types plays very important and central role in the classical information theory. With it the central theorems can be easily and fast proved. This work try to generalise the main ideas of the method of types, and prove one of the central theorems - Reaching the Holevo capacity - in quantum environment.

The main problem is: Suppose that we want to send information with quantum's. This problem is relevant, because 1) quantum computers would prefer this way of communication 2) The miniaturisation in the Information Technologies can lead to these type of problem. The problem can be formalised as follows: We code our classical message to quantum sequences states from $(\omega_1, \omega_2, \ldots, \omega_l)$. We suppose that there is a unique non-reversible quantum transformation $\mathcal{E}(\cdot)$ - quantum channel - which acts on every of these quantum's. The question is how many bits of information can be transmitted by one quantum. A theorem stated by Gordon and Levitin, proved by Holevo [4], gives an upper bound to the amount of information that can be communicated. If the sender codes his information to quantum states with density matrix ρ_i with a priori probabilities p_l then the communicated information cannot be bigger than

$$S(\sum_{i=1}^{l} p_i \rho_i) - \sum_{i=1}^{l} p_i S(\rho_i) \tag{1}$$

where $S(\cdot)$ is the von Neumann entropy. If the outcome of the channel is $\rho_l = \mathcal{E}(\omega_l)$ this gives an upper bound. So the problem is to show, that this bound can be reached.

At the end, our result is stronger than the work of Holevo [5] or Schumacher and Westmoreland [3], because we will show that the decoding can be done by von Neumann measurement, not only with POVM (We doesn't use "Pretty Good Measurement" as in [2] or [5]). Moreover, we show two use of the von Neumann measurement.

The first use is that the procedure can be generalised to finite compound channel, that means, we can create an optimal coding scheme to work not only

1

with one quantum channel, but with finite many. Definition and capacity is in section 4

The other use is that we can translate decoding of classical information to decoding of classical quantum information, and with the von Neumann measurement the time of the measuring procedure can be extremely shorten, which means that classical information can be decoded by a quantum apparatus in linear! time (This is a strong result, the best codes which reach the Shannon's bound needs $n^{\log_2(n)}$ time to decode).

This work base notation is borrowed from the work of Schumacher and Westmoreland [3], but the base ideas of the proofs, comes from the classical theory e.g. [6].

2 Notations and basic lemmas

Let $\mathcal{E}(\,\cdot\,)$ be a given quantum channel. Assume that $\omega_1, \omega_2, \ldots, \omega_l = \omega_1^l$ are input density matrices, with same dimension d and $P = (p_1, p_2, \ldots, p_n)$ is a probability distribution such that they maximise the Holevo quantity

$$\chi(\mathcal{E}, P, \omega_1^l) = S(\mathcal{E}(\omega)) - \sum_{i=1}^{l} p_i S(\mathcal{E}(\omega_i)), \qquad (2)$$

where $\omega = \sum_{i=1}^{l} p_i \omega_i$. Denote the possible outputs of the quantum channel by $\rho_i = \mathcal{E}(\omega_i), \rho = \mathcal{E}(\omega)$, these are represented by $d \times d$ density matrices.

Fix n, the length of the (quantum) codewords. We generate randomly $M = 2^{nR}$ piece codewords of length n with probability distribution P. We denote these randomly generated sequence by $\alpha_i, \quad 1 \le i \le M$. If a statement is true all of the index $1 \le i \le M$, then we will say that it is true for α. The j-th symbol of α will be denoted by $\alpha(j)$.

For all sequence we can define a quantum sequence as follows:

$$\rho_\alpha = \rho_{\alpha(1)} \otimes \rho_{\alpha(2)} \otimes \cdots \otimes \rho_{\alpha(n)} \qquad (3)$$

We will denote by $S(\rho|\alpha)$ the quantity $\sum_{j=1}^{l} p_j S(\rho_j)$, because $nS(\rho|\alpha)$ is the expected value of the Neumann entropy of the quantum sequence if we know which randomly generated sequence (α) was sent (so we know which basis to use), while $S(\rho)$ is the von Neumann entropy of the sequence if we do not know which message was sent. So for fixed P and ω the Holevo capacity becomes

$$\chi = S(\rho) - S(\rho|\alpha) \qquad (4)$$

Which resembles the Shannon capacity

$$C = H(Y) - H(Y|X) \qquad (5)$$

where Y is the output random variable and X is an input random random variable of the channel.

For a fixed $\varepsilon > 0$, a sequence α is called ε-typical with respect to P if

$$2^{-n(H(P)+\varepsilon)} \le P^n(\alpha) \le 2^{-n(H(P)-\varepsilon)}, \tag{6}$$

where $H(P)$ is the Shannon entropy of P. We know from the law of large numbers that, if n is large enough then the probability

$$\text{Prob}\,(\alpha \text{ is typical}) \ge 1 - \varepsilon \tag{7}$$

(because α was generated by distribution P), see [6].

Let the spectral decomposition of $\mathcal{E}(\rho_\alpha) = \sum_{k=1}^{d^n} \lambda_{\alpha,k} |s_{\alpha,k}\rangle\langle s_{\alpha,k}|$. Because ρ_α is a tensor product, the eigenvectors are tensor products of eigenvectors of the $\rho_{\alpha(1)}, \rho_{\alpha(2)}, \dots, etc.$. So a measurement in the eigenbasis can be represented by a sequence, from numbers $\{1, 2, \dots, d\}$ where the i-th term gives that what we would measure if we measure $\rho_{\alpha(i)}$ in its eigenbasis. Denote this correspondence by $s : \{1, 2, \dots d^n\} \to \{1, 2, \dots d\}^n$, note that $\lambda_k = \lambda_{s(k)_1} \lambda_{s(k)_2} \cdots \lambda_{s(k)_n}$. An eigenvector $|s_{\alpha,k}\rangle$ is δ-typical if the above defined distribution $(\lambda_{s(k)_1} \lambda_{s(k)_2} \cdots)$ is conditionally typical to the sequence α (see [6])

$$- n(S(\rho|\alpha) + \delta) \le \log \lambda_{\alpha,k} \le -n(S(\rho|\alpha) - \delta) \tag{8}$$

Note that all exponent and logarithm are base of 2 across of this article. The above definition means that if we define the typical projection as

$$Pi_\alpha = \sum_{k:s_{\alpha,k}\text{is typical}} |s_{\alpha,k}\rangle\langle s_{\alpha,k}| \tag{9}$$

, then

$$dim(\Pi_\alpha) \le 2^{nS(\rho|\alpha)+\delta} \tag{10}$$

while

$$1 = \sum_{k=1}^{d^n} \lambda_{\alpha,k} \ge \sum_{k:s_{\alpha,k}\text{is typical}} \lambda_{\alpha,k} \ge \sum_{k:s_{\alpha,k}\text{is typical}} 2^{-n(S(\rho|\alpha)+\delta)} \tag{11}$$

$$2^{n(S(\rho|\alpha)+\delta)} \ge \sum_{k:s_{\alpha,k}\text{is typical}} 1 = dim(\Pi_\alpha) \tag{12}$$

The pair (α, k) is δ typical if

$$- n\Big(H(P) + S(\rho|\alpha) + \delta\Big) \le \log P^n(\alpha)\lambda_{\alpha,k} \le -n\Big(H(P) + S(\rho|\alpha) - \delta\Big). \tag{13}$$

Let the distribution of $(\alpha, s(k))$ be denoted by

$$P_{\alpha,s(k)} = \Pi_{i=1}^n P(\alpha(i))\lambda_{\alpha(i),s(k)_i}, \tag{14}$$

it can be seen that this is a probability of independent, identically distributed random variables. The (Shannon) entropy of this distribution is

$$H(P) + \sum_{i=1}^l p_i S(\rho_i). \tag{15}$$

3

So for these pair of random variables, the law of large numbers also true, so then by summing the probability of all typical pair we also get a greater number than $1 - \varepsilon$ if n is large enough. Suppose that the indexing of the eigenvalues is such that the typical eigenvalues are the first $d(\alpha)$ indices. Then

$$\sum_\alpha p_\alpha \sum_{i=1}^{d(\alpha)} \lambda_{\alpha,i} \geq 1 - 2\varepsilon \tag{16}$$

because from the sum we only left the atypical α (which probability is smaller that ε) and atypical α, k pairs (which probability is also smaller that ε) From this we know, that for $\tilde{\rho} = \sum_\alpha p_\alpha \sum_{i=1}^{d(\alpha)} \lambda_{\alpha,i} |s_{\alpha,i}\rangle\langle s_{\alpha,i}|$ it is true that $\operatorname{Tr} \tilde{\rho} \geq 1 - 2\varepsilon$. Define $\tilde{\rho}_\alpha$ as

$$\tilde{\rho}_\alpha = \Pi_\alpha \rho_\alpha \Pi_\alpha \tag{17}$$

Remark 1 $\tilde{\rho} \leq \rho^{\otimes n}$ and $\tilde{\rho}_\alpha \leq \rho_\alpha$, and $\mathbb{E}(\tilde{\rho}_\alpha) = \tilde{\rho}$

Lemma 1 For every $n \in \mathbb{N}$

$$\mathbb{E}[\rho_\alpha] = \rho^{\otimes n} \tag{18}$$

Proof: Total induction on n For $n = 1$ the equivalence is true by the definition of ρ. Suppose that for $n = k-1$ the statement is true, then For $n = k$ let denote α' an arbitrary $k-1$ length sequence then for every α k length sequence can be written as $\alpha = (\alpha', i), \quad 1 \leq i \leq l$. Then

$$\mathbb{E}[\rho_\alpha] = \sum_\alpha p^k(\alpha)\rho_\alpha = \sum_{\alpha'} \sum_{i=1}^l p^{k-1}(\alpha')p(i)\rho_{\alpha'} \otimes \rho_i = \tag{19}$$

$$= \sum_{\alpha'} p^{k-1}(\alpha') \sum_{i=1}^l p_i \rho_{\alpha'} \otimes \rho_i = \sum_{\alpha'} p^{k-1}(\alpha')\rho_{\alpha'} \otimes \rho \tag{20}$$

but we know that for $n = k-1$ the statement is true, so

$$\mathbb{E}[\rho_\alpha] = \rho^{\otimes k-1} \otimes \rho = \rho^{\otimes k} \tag{21}$$

So the statement is true for all $n \in \mathbb{N}$. ∎

If we have a projection, like Π_α then we can define a subspace which this projection projects to $\pi_\alpha = Im(\Pi_\alpha)$. And vice versa, if we define a subspace π_α, then we can define an orthogonal projection which project to this subspace Π_α. This will be done throughout the paper by denoting with the same letter, indices the lowercase denotes the subspace the uppercase denotes the projections.

Consider the lattice of the projections. For 2 projection P_1 and P_2 denote $P_1 \vee P_2$ the projection which is the result of the $\vee$ operation in the net of projections (this means that $P_1 \vee P_2$ is the projection which project to the subspace spanned by the range of P_1 and P_2). Similarly meaning has $P_1 \wedge P_2$ ($P_1 \wedge P_2$ s the projection which projects to the subspace which is the section of the range of P_1 and P_2).

Lemma 2 *For every density matrices ρ, σ and for every $\varepsilon > 0$ there exist a projection Π, with properties*

$$\mathrm{Tr}(\Pi \rho^{\otimes n} \Pi) \geq 1 - \varepsilon \tag{22}$$

$$\|\Pi \rho^{\otimes n} \Pi\| \leq 2^{-n(S(\rho) - \varepsilon)} \tag{23}$$

$$\mathrm{Tr}(\Pi \sigma^{\otimes n} \Pi) \leq 2^{-n(D(\rho|\sigma) - \varepsilon)} \tag{24}$$

if $n > N(\rho, \sigma, \varepsilon)$, where $D(\rho|\sigma)$ denotes the quantum relative entropy of ρ, σ.

for proof see the appendix.

Remark 2 *Note that, Π does not depend on α_i the randomly chosen sequence, but $\tilde{\Pi}_{i-}$ does (across the article if an amount depend on the randomly generated sequence, then it will denoted by an α_i in the index or in the argument). This means that $\mathbb{E}(\Pi \rho_{\alpha_i} \Pi) = \Pi \rho^{\otimes n} \Pi$ while $\mathbb{E}(\rho_{\alpha_i}) = \rho^{\otimes n}$, but $\mathbb{E}(\tilde{\Pi}_{i-} \rho_{\alpha_i} \tilde{\Pi}_{i-}) \neq \tilde{\Pi}_{i-} \rho^{\otimes n} \tilde{\Pi}_{i-}$.*

3 Reaching the Holevo bound with von Neumann measurement

3.1 Coding/Decoding

First we generate $M = 2^{nR}$ random codewords with distribution P. These codewords are denoted by $\alpha_1, \alpha_2, \ldots, \alpha_M$ and both the sender, and receiver are familiar with them. From these we generate quantum codewords.

A quantum codeword is a tensor product density defined in the following way: If $\alpha_i(j) = k$, this is the jth symbol of the ith codeword, then the jth density of the ith tensor product is ω_k. The coding is as usual, we choose uniformly from the message set $\mathcal{M}$ whose size is M − suppose this chosen message is i −, and we send (or generate) the above defined quantum codeword for this message $\omega_{\alpha_i(1)}, \omega_{\alpha_i(2)}, \ldots, \omega_{\alpha_i(n)}$. The quantum sequences go through the channel, the receiver gets the quantum sequence $\rho_{\alpha_i(1)}, \rho_{\alpha_i(2)}, \ldots, \rho_{\alpha_i(n)}$ where $\rho_{\alpha_i(1)} = \mathcal{E}(\omega_{\alpha_i(1)})$.

Now we define a decoding algorithm which is nothing else than a POVM (von Neumann measurement). If the typical subspaces of ρ_{α_i} were orthogonal to each other, then - there would be no problem - we could make our POVM from Π_{α_i} (where Π_{α_i} defined in (9)). However, this is not the case in general, and we have to orthogonalize them. We do this by a method, very similar to the Gram-Schmidt orthogonalisation method. In the first typical subspace let

$$\tilde{\pi}_1 = span\{|\tilde{s}_{\alpha_1, k}\rangle : |\tilde{s}_{\alpha_1, k}\rangle = \Pi |s_{\alpha_1, k}\rangle, \ |s_{\alpha_1, k}\rangle \in \pi_{\alpha_1}\} \tag{25}$$

where $span\{\}$ means the subspace spanned by the the vectors in the curly bracket.

And for the i-th typical projections let

$$\tilde{\pi}_i = span \left\{ |\tilde{s}_{\alpha_1,k}\rangle : |\tilde{s}_{\alpha_1,k}\rangle = \Pi|s_{\alpha_i,k}\rangle - \sum_{j=1}^{i-1} \tilde{\Pi}_j \Pi|s_{\alpha_i,k}\rangle, \ |s_{\alpha_i,k}\rangle \in \pi_{\alpha_i} \right\} \quad (26)$$

(for defining subspaces to projections, and vice versa see Section 2)

At the end of the procedure for all s,t $\tilde{\Pi}_s$ is orthogonal to $\tilde{\Pi}_t$ $s \neq t$. We prove this, by total induction on s,t. Suppose that $s < t$. We see that for $s = 1, t = 2$ this is true because

$$\tilde{\Pi}_1|\tilde{s}_{\alpha_2}\rangle = \tilde{\Pi}_1\Pi|s_{\alpha_2}\rangle - \tilde{\Pi}_1\tilde{\Pi}_1\Pi|s_{\alpha_2}\rangle = \tilde{\Pi}_1\Pi|s_{\alpha_2}\rangle - \tilde{\Pi}_1\Pi|s_{\alpha_2}\rangle = 0 \quad (27)$$

so $\tilde{\pi}_2 \in Ker(\tilde{\Pi}_1)$. Suppose that, for all pairs s',t', where $s' < s, t' < t$ the statement is true. then

$$\tilde{\Pi}_s|\tilde{s}_{\alpha_t}\rangle = \tilde{\Pi}_s\Pi|s_{\alpha_t}\rangle - \tilde{\Pi}_s \sum_{t'=1}^{t-1} \tilde{\Pi}_{t'}\Pi|s_{\alpha_2}\rangle = \tilde{\Pi}_s\Pi|s_{\alpha_t}\rangle - \tilde{\Pi}_s\Pi|s_{\alpha_t}\rangle = \underline{0} \quad (28)$$

because $\tilde{\Pi}_s\tilde{\Pi}_{t'} = 0$ for $t' \neq s$ by the induction assumption. Moreover the $\tilde{\Pi}_i$ and Π are commutable operators (because every vector which spans the subspace of $\tilde{\Pi}_i$ projects onto is a member of π).

Our POVM (or our Von-Neumann measure) states from $\tilde{\Pi}_i$, plus we can make it complete adding an element of the POVM (labelled "error") on the remaining orthogonal subspace, if necessary, call these projection as $\tilde{\Pi}_{M+1}$. For index i, we can define the typical subspace of lesser indices as

$$\tilde{\Pi}_{i-} = \sum_{j=1}^{i-1} \tilde{\Pi}_j$$

3.2 The error probability of Decoding

Now we show that the error probability is going to 0 if n the block length goes to infinity.

Theorem 1 *With these scheme, if $R < \chi(\rho)$ then for any γ we can give a number N such that if the length of the quantum codeword n is longer than this number $n > N$ then average error probability is smaller than γ, provided that the blocklength n is greater than $n_0(R,\gamma)$*

To simplify the proof we use the following lemma.

Lemma 3 *The length of the projection of $|s_{\alpha_i,j}\rangle$ to $\tilde{\Pi}_i$ (element of our measuring POVM) can be underestimated - by two component where one is the projection to Π, and the other is orthogonal to the typical subspace of the lesser indices ($\tilde{\Pi}_{i-}$) - as follows*

$$\mathrm{Tr}(\tilde{\Pi}_i \rho_{\alpha_i} \tilde{\Pi}_i) \geq \left(\mathrm{Tr}(\Pi\tilde{\rho}_{\alpha_i}\Pi) - \mathrm{Tr}(\tilde{\Pi}_{i-}\Pi\tilde{\rho}_{\alpha_i}\Pi\tilde{\Pi}_{i-1}) \right)^2 \quad (29)$$

(For definition of $\tilde{\rho}_{\alpha_i}$ see (17))

6

Proof:

$$\text{Tr}(\tilde{\Pi}_i \rho_{\alpha_i} \tilde{\Pi}_i) = \text{Tr}(\tilde{\Pi}_i \rho_{\alpha_i}) = \text{Tr}(\tilde{\Pi}_i \sum_{j=1}^{d} \lambda_{\alpha_i,j} |s_{\alpha_i,j}\rangle\langle s_{\alpha_i,j}|) = \tag{30}$$

Now we decompose the projection $\tilde{\Pi}_i$ depend on j as follows

$$\tilde{\Pi}_i = a_j^2 |\tilde{s}_{\alpha_i,j}\rangle\langle \tilde{s}_{\alpha_i,j}| + \hat{\Pi}_i \tag{31}$$

where $a_j > 1$ is the reciprocal of the length of $|\tilde{s}_{\alpha_i,j}\rangle = \Pi|s_{\alpha_i,k}\rangle - \sum_{j=1}^{i-1} \tilde{\Pi}_j \Pi |s_{\alpha_i,k}\rangle$ (length of $|\tilde{s}_{\alpha_i,j}\rangle$ is smaller than 1 because it is an orthogonal component of a unit length vector $|s_{\alpha_i,j}\rangle$). So we continue the first row by

$$= \sum_{j=1}^{d} \lambda_j \, \text{Tr}(a_j^2 |\tilde{s}_{\alpha_i,j}\rangle\langle \tilde{s}_{\alpha_i,j}||s_{\alpha_i,j}\rangle\langle s_{\alpha_i,j}| + \hat{\Pi}_i |s_{\alpha_i,j}\rangle\langle s_{\alpha_i,j}|) = \tag{32}$$

$$\sum_{j=1}^{d} \lambda_j \, \text{Tr}(a_j^2 |\tilde{s}_{\alpha_i,j}\rangle\langle \tilde{s}_{\alpha_i,j}||s_{\alpha_i,j}\rangle\langle s_{\alpha_i,j}|) + \sum_{j=1}^{d} \lambda_j \, \text{Tr}(\langle s_{\alpha_i,j}|\hat{\Pi}_i|s_{\alpha_i,j}\rangle) \geq \tag{33}$$

$$\geq \sum_{j=1}^{d(\alpha_i)} \lambda_j \, \text{Tr}(|\tilde{s}_{\alpha_i,j}\rangle\langle \tilde{s}_{\alpha_i,j}||s_{\alpha_i,j}\rangle\langle s_{\alpha_i,j}|) = \tag{34}$$

$$= \sum_{j=1}^{d(\alpha_i)} \lambda_j \, \text{Tr}((\Pi|s_{\alpha_i,j}\rangle - \tilde{\Pi}_{i-}\Pi|s_{\alpha,j}\rangle)(\langle s_{\alpha_i,j}|\Pi - \langle s_{\alpha_i,j}|\Pi \tilde{\Pi}_{i-})|s_{\alpha_i,j}\rangle\langle s_{\alpha_i,j}|) = \tag{35}$$

$$= \sum_{j=1}^{d(\alpha_i)} \lambda_j \left(\langle s_{\alpha_i,j}|(\Pi|s_{\alpha_i,j}\rangle - \Pi_{i-}\Pi|s_{\alpha,j}\rangle)\right)^2 \tag{36}$$

Where (34) comes from the fact, that $a_j > 1$ so $a_j^2 > 1$. From the Jensen's inequality

$$\geq \left(\sum_{j=1}^{d(\alpha_i)} \lambda_j \langle s_{\alpha_i,j}|(\Pi|s_{\alpha_i,j}\rangle - \tilde{\Pi}_{i-}\Pi|s_{\alpha,j}\rangle)\right)^2 = \tag{37}$$

$$= \left(\sum_{j=1}^{d(\alpha_i)} \lambda_j \langle s_{\alpha_i,j}|\Pi|s_{\alpha_i,j}\rangle - \sum_{j=1}^{d(\alpha_i)} \lambda_j \langle s_{\alpha_i,j}|\tilde{\Pi}_{i-}\Pi|s_{\alpha_i,j}\rangle\right)^2 = \tag{38}$$

$$= \left(\sum_{j=1}^{d(\alpha_i)} \lambda_j \, \text{Tr}(\Pi|s_{\alpha_i,j}\rangle\langle s_{\alpha_i,j}| - \sum_{j=1}^{d(\alpha_i)} \lambda_j \, \text{Tr}(\tilde{\Pi}_{i-}\Pi|s_{\alpha_i,j}\rangle\langle s_{\alpha_i,j}|)\right)^2 = \tag{39}$$

$$= \left(\text{Tr}(\Pi\tilde{\rho}_{\alpha_i}) - \text{Tr}(\tilde{\Pi}_{i-}\Pi\tilde{\rho}_{\alpha_i})\right)^2 \geq \left(\text{Tr}(\Pi\tilde{\rho}_{\alpha_i}\Pi) - \text{Tr}(\tilde{\Pi}_{i-}\Pi\tilde{\rho}_{\alpha_i}\Pi\tilde{\Pi}_{i-})\right)^2 \tag{40}$$

■

Now see the proof:

Proof of Theorem 1: Let ε is such small that $R + \varepsilon < \chi(\rho)$, and $8\varepsilon \leq \gamma$ and let n such large that all the Lemmas and definitions from Section 2 with ε are true.

$$P_e = \mathbb{E}[1 - \frac{1}{M}\sum_{i=1}^{M}\mathrm{Tr}(\tilde{\Pi}_i \rho_{\alpha_i} \tilde{\Pi}_i)] \leq \tag{41}$$

$$\leq \frac{1}{M}\sum_{i=1}^{M}\mathbb{E}[1 - \mathrm{Tr}(\tilde{\Pi}_i \rho_{\alpha_i} \tilde{\Pi}_i)] \leq \tag{42}$$

$$\leq \frac{1}{M}\sum_{i=1}^{M}\mathbb{E}[1 - \mathrm{Tr}(\tilde{\Pi}_i \rho_{\alpha_1} \tilde{\Pi}_i)] \tag{43}$$

We use the Lemma

$$P_e \leq \frac{1}{M}\sum_{i=1}^{M} 1 - \mathbb{E}[\mathrm{Tr}(\Pi\tilde{\rho}_{\alpha_i}\Pi - \tilde{\Pi}_{i-}\Pi\tilde{\rho}_{\alpha_i}\Pi\tilde{\Pi}_{i-})]^2 \tag{44}$$

We use the Jensen's inequality

$$P_e \leq \frac{1}{M}\sum_{i=1}^{M} 1 - [\mathbb{E}\,\mathrm{Tr}(\Pi\tilde{\rho}_{\alpha_i}\Pi - \tilde{\Pi}_{i-}\Pi\tilde{\rho}_{\alpha_i}\Pi\tilde{\Pi}_{i-})]^2 = \tag{45}$$

$$= \frac{1}{M}\sum_{i=1}^{M} 1 - [\mathrm{Tr}(\Pi\mathbb{E}(\tilde{\rho}_{\alpha_i})\Pi - \tilde{\Pi}_{i-}\Pi\mathbb{E}(\tilde{\rho}_{\alpha_i})Pi\tilde{\Pi}_{i-})]^2 = \tag{46}$$

$$\frac{1}{M}\sum_{i=1}^{M} 1 - [\mathrm{Tr}(\Pi\tilde{\rho}\Pi) - \mathbb{E}(\mathrm{Tr}(\tilde{\Pi}_{i-}\Pi\tilde{\rho}_{\alpha_i}\Pi\tilde{\Pi}_{i-}))]^2 \tag{47}$$

From (16), and remark 1, we know that $\mathrm{Tr}(\tilde{\rho}) \geq 1 - 2\varepsilon$ and $\rho \geq \tilde{\rho}$ so, for $\Delta = \rho - \tilde{\rho}$, $\mathrm{Tr}(\Delta) \leq 2\varepsilon$. With this

$$\mathrm{Tr}(\Pi\tilde{\rho}\Pi) = \mathrm{Tr}(\Pi\rho\Pi) - \mathrm{Tr}(\Pi(\rho - \tilde{\rho})\Pi) \geq 1 - \varepsilon - \mathrm{Tr}(\Delta) \tag{48}$$

So the error probability

$$P_e \leq \frac{1}{M}\sum_{i=1}^{M} 1 - [1 - 3\varepsilon - \mathbb{E}(\mathrm{Tr}(\tilde{\Pi}_{i-}\Pi\tilde{\rho}_{\alpha_i}\Pi\tilde{\Pi}_{i-}))]^2 \leq \tag{49}$$

$$\leq \frac{1}{M}\sum_{i=1}^{M} 1 - [1 - 3\varepsilon - \mathbb{E}(\mathrm{Tr}(\tilde{\Pi}_{i-}\Pi\tilde{\rho}_{\alpha_i}\Pi\tilde{\Pi}_{i-}))]^2 \tag{50}$$

8

Now we analyse the last term

$$\mathbb{E}(\mathrm{Tr}(\tilde{\Pi}_{i-}\Pi\tilde{\rho}_{\alpha_i}\Pi\tilde{\Pi}_{i-})) = \mathbb{E}(\mathrm{Tr}(\tilde{\Pi}_{i-}\Pi\tilde{\rho}_{\alpha_i}\Pi)) = \mathbb{E}(\mathrm{Tr}(\sum_{j=1}^{dim(\tilde{\Pi}_{i-})} |b_j\rangle\langle b_j|\Pi\tilde{\rho}_{\alpha_i}\Pi)) = \tag{51}$$

$$\mathbb{E}(\sum_{i=1}^{dim(\tilde{\Pi}_{i-})} \langle b_j|\Pi\tilde{\rho}_{\alpha_i}\Pi|b_j\rangle) \tag{52}$$

Now we have to evaluate $dim(\tilde{\Pi}_{i-})$. Because of (10)

$$dim(\tilde{\Pi}_{i-}) \leq \sum_{j=2}^{M} dim(\Pi_{\alpha_j}) \leq M2^{nS(\rho|\alpha)} = 2^{n(R+S(\rho|\alpha))} \tag{53}$$

$$\mathbb{E}(\mathrm{Tr}(\tilde{\Pi}_{i-}\Pi\tilde{\rho}_{\alpha_i}\Pi\tilde{\Pi}_{i-})) \leq 2^{n(R+S(\rho|\alpha))} \max_j \mathbb{E}(\langle b_j|\Pi\tilde{\rho}_{\alpha_i}\Pi|b_j\rangle) = \tag{54}$$

$$2^{n(R+S(\rho|\alpha))} \max_j \langle b_j|\Pi\tilde{\rho}\Pi|b_j\rangle \leq 2^{n(R+S(\rho|\alpha))} \max_j \langle b_j|\Pi\rho\Pi|b_j\rangle \tag{55}$$

Because b_j's are unit vectors, from (23) we know that $\max_j\langle b_j|\Pi\rho\Pi|b_j\rangle \leq 2^{-n(S(\rho)-\varepsilon)}$, so

$$\mathbb{E}(\mathrm{Tr}(\tilde{\Pi}_{i-}\Pi\tilde{\rho}_{\alpha_i}\Pi\tilde{\Pi}_{i-})) \leq 2^{n(R+S(\rho|\alpha))}2^{-n(S(\rho)-\varepsilon)} \leq 2^{-n(S(\rho)-S(\rho|\alpha)-R-\varepsilon)} \tag{56}$$

Now we can see that $S(\rho) - S(\rho|\alpha) = \chi(\rho)$, and we assumed that $R + \varepsilon < \chi(\rho)$, so the exponent is negative. If n is large enough then the whole expression is less than ε.

$$\mathbb{E}(\mathrm{Tr}(\tilde{\Pi}_{i-}\Pi\tilde{\rho}_{\alpha_i}\Pi\tilde{\Pi}_{i-})) \leq \varepsilon \tag{57}$$

So the error probability is smaller than

$$P_e \leq 1 - [1 - 4\varepsilon]^2 = 8\varepsilon - 16\varepsilon^2 \leq 8\varepsilon \tag{58}$$

∎

4 I. Use:
Coding for finite compound channel

First we give a definition of the compound channel. Suppose there is a given set of channels $\mathcal{S}$. We want a predefined coding scheme to code our message with the following disturbing effect: suppose there is an enemy who chose one channel from the set after we generated our quantum codeword. Now our quantum

codeword are send through the chosen channel (so all the quantum has the same effect but we cannot say which).

This is a more realistic model than a simple quantum channel, we know what effects can destroy our quantum codewords,but we cannot know at the present moment, which effect is active. This compound channel is a finite compound channel, if the set $\mathcal{S}$ is finite. Define:

$$\chi(\mathcal{S}, P, \omega_1^l) \stackrel{\circ}{=} \min_{\mathcal{E} \in \mathcal{S}} \chi(\mathcal{E}, P, \omega_1^l) \tag{59}$$

$$\chi(\mathcal{S}) \stackrel{\circ}{=} \max_{P, \omega_1^l} \chi(\mathcal{S}, P, \omega_1^l) \tag{60}$$

Theorem 2 *The classical capacity C of the finite compound channel is $\chi(\mathcal{S})$, This means, if $R < \chi(\mathcal{S})$ then for any γ there exist a number $N(\gamma, R, |\mathcal{S}|)$ that if the length of the quantum codeword n is larger than that number $n > N(\gamma)$ then the error probability is smaller than γ.*

Proof: First we can assume that $|\mathcal{S}| > 2$, because for $|\mathcal{S}| = 1$, the proof is the same as theorem 1. The inequality $C \leq \chi(\mathcal{S})$ is simple. Because in [4] was shown, that a quantum system can carry $\chi(\mathcal{E}, P, \omega_1^l)$ bit information if the sender code his message to quantum states $\mathcal{E}(\omega_i)$ with a priori probabilities P. If the sender has no knowledge, which channel is being used, then the sender can only codes his message to states ω_i with a priori distribution P. Then if his enemy chose the worst channel for these schema, the carried information cannot be larger than

$$\max_{P, \omega_1^l} \min_{\mathcal{E} \in \mathcal{S}} \chi(\mathcal{E}, P, \omega_1^l) = \chi(\mathcal{S}).$$

To prove that $C \geq \chi(\mathcal{S})$ we show a coding scheme which rate can reach the capacity.

4.1 Coding for finite compound channel

Let $|\mathcal{S}| = a$, $\mathcal{S} = \mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_a$ fixed. Let P, and $\omega_1, \omega_2, \ldots, \omega_l$ be the probability distribution and quantum's that reach the maximum in (60). The sender codes his message to randomly generated codeword as in section 3. The codeword goes into the quantum channel $\mathcal{E}_o$ $o \in \{1, 2, \ldots, a\}$ which was chosen by our enemy. Denote $\omega = \sum_{j=1}^l p_l \omega_l$ the input mixed state, and denote $\rho^k = \mathcal{E}_k^{\otimes n}(\omega^{\otimes n})$, $1 \leq k \leq a$ the possible mixed output of the channel. Similarly denote $\rho_{\alpha_i}^k = \mathcal{E}_k(\rho_{\alpha_i})$ the possible output of the i-th quantum codeword. To simplify our proof we can assume that, the order of the set of quantum channels is such that, the first $\bar{a}$ $\rho^k, k \in \{1, 2, \ldots, \bar{a}\}$ is different.

4.2 Decoding for finite compound channel

Decoding is done by two steps: In first step we can detect which mixed state we have. In the second step we detect the message.

See the first step. Let denote $\bar{a}$ the number of the different output mixed states. To distinguish the output mixed states we will use our Lemma 2. Let ε be such small that $\varepsilon < \min_{i \neq j 1 \leq i,j \leq \bar{a}} \frac{D(\rho^i \| \rho^j)}{2}$, $\varepsilon \leq \frac{\gamma}{8a}$ and $R + \varepsilon < \chi(\mathcal{S})$. Then for an n large enough for every $i,j : i \neq j 1 \leq i,j \leq \bar{a}$ Lemma 2 is true, with $\rho = \rho^l, \sigma = \rho^k$ and get $\Pi^{i,j}$ (if n is greater than $\max(N(\rho^l, \rho^k, \varepsilon))$). Now we make a typical projection for all i as follows:

$$\Pi^i = \bigwedge_{1 \leq j \leq \bar{a}} \Pi^{i,j} \tag{61}$$

See, that

$$\mathrm{Tr}(\Pi^i (\rho^i)^{\otimes n} \Pi^i) = 1 - \mathrm{Tr}((\Pi^i)^c (\rho_i)^{\otimes n} (\Pi^i)^c) \geq \tag{62}$$

$$\geq 1 - \sum_{j \neq i, 1 \leq j \leq \bar{a}} \mathrm{Tr}((\Pi^{i,j})^c (\rho_i)^{\otimes n} (\Pi^{i,j})^c) \geq 1 - a\varepsilon \tag{63}$$

$$\|\Pi^i (\rho^i)^{\otimes n} \Pi^i\| \leq 2^{-n(S(\rho_i) - \varepsilon)} \tag{64}$$

$$\mathrm{Tr}(\Pi^i (\rho^j)^{\otimes n} \Pi^i) \leq 2^{-n(D(\rho^i \| \rho^j) - \varepsilon)} \tag{65}$$

for all $1 \leq i,j \leq a$.

We detect the mixed state as follows: First we give a sequence of measures. Our first POVM measure states from $\Pi^1, (\Pi^1)^c$ if we measure Π^1 we know, that our mixed state was ρ^1 so we stop, if we measured $(\Pi^1)^c$ then we measure again. Second POVM measure states from $\Pi^2, (\Pi^2)^c$, etc. With this we can differentiate the possible $\bar{a}$ type of our mixed state.

We suppose that our enemy chosen the channel $\mathcal{E}_o$ which generate the k-th mixed state (this means $\mathcal{E}_l(\omega) = \mathcal{E}_o(\omega)$), if $k \neq 1$ then the first measure is good, if we measure the second projection. This probability is

$$\mathrm{Tr}((\Pi^1)^c \rho^k_{\alpha_i} (\Pi^1)^c) \tag{66}$$

and our quantum states will be

$$\frac{(\Pi^1)^c \rho^k_{\alpha_i} (\Pi^1)^c}{\mathrm{Tr}((\Pi^1)^c \rho^k_{\alpha_i} (\Pi^1)^c)} \tag{67}$$

Next if $k \neq 2$ then the next measure is good if we measure the second projection this probability is

$$\mathrm{Tr}((\Pi^2)^c \frac{(\Pi^1)^c \rho^k_{\alpha_i} (\Pi^1)^c}{\mathrm{Tr}((\Pi^1)^c \rho^k_{\alpha_i} (\Pi^1)^c)} (\Pi^2)^c) \tag{68}$$

and our state becomes

$$\frac{(\Pi^2)^c \frac{(\Pi^1)^c \rho^k_{\alpha_i} (\Pi^1)^c}{\mathrm{Tr}((\Pi^1)^c \rho^k_{\alpha_i} (\Pi^1)^c)} (\Pi^2)^c}{\mathrm{Tr}((\Pi^2)^c \frac{(\Pi^1)^c \rho^k_{\alpha_i} (\Pi^1)^c}{\mathrm{Tr}((\Pi^1)^c \rho^k_{\alpha_i} (\Pi^1)^c)} (\Pi^2)^c)} = \frac{(\Pi^2)^c (\Pi^1)^c \rho^k_{\alpha_i} (\Pi^1)^c (\Pi^2)^c}{\mathrm{Tr}(\Pi^2)^c (\Pi^1)^c \rho^k_{\alpha_i} (\Pi^1)^c (\Pi^2)^c)} \tag{69}$$

11

And the probability, that we don't made error through the first, and the second step is

$$\text{Tr}\left(\frac{(\Pi^2)^c(\Pi^1)^c\rho_{\alpha_i}^k(\Pi^1)^c(\Pi^2)^c}{\text{Tr}((\Pi^1)^c\rho_{\alpha_i}^k(\Pi^1)^c)}\right)\text{Tr}((\Pi^1)^c\rho_{\alpha_i}^k(\Pi^1)^c) = \tag{70}$$

$$\text{Tr}((\Pi^2)^c(\Pi^1)^c\rho_{\alpha_i}^k(\Pi^1)^c(\Pi^2)^c) \tag{71}$$

From this we can see that, if $\mathcal{E}_o(\omega) = \mathcal{E}_k(\omega)$ then the probability that we detect our mixed state correctly is:

$$\mathbb{E}[\text{Tr}(\Pi^k(\Pi^{k-1})^c\dots(\Pi^2)^c(\Pi^1)^c\rho_{\alpha_i}^k(\Pi^1)^c(\Pi^2)^c\dots(\Pi^{k-1})^c\Pi^k)] \tag{72}$$

Let $\delta = a\varepsilon$. From Lemma 1, and from the definition of Π^k we know

$$\mathbb{E}[\text{Tr}(\Pi^k\rho_{\alpha_i}^k\Pi^k)] = \text{Tr}(\Pi^k\rho^k\Pi^k) \geq 1 - \delta \tag{73}$$

Moreover

$$1 - \delta \leq \text{Tr}(\Pi^k\rho^k\Pi^k) \leq \text{Tr}(\Pi^k(\Pi^1)^c\rho^k(\Pi^1)^c\Pi^k) + \text{Tr}(\Pi^k\Pi^1\rho^k\Pi^1\Pi^k) \leq \tag{74}$$

$$\leq \text{Tr}(\Pi^k(\Pi^1)^c\rho^k(\Pi^1)^c\Pi^k) + \text{Tr}(\Pi^1\rho^k\Pi^1) \tag{75}$$

$$\leq \text{Tr}(\Pi^k(\Pi^1)^c\rho^k(\Pi^1)^c\Pi^k) + 2^{-nD(\rho^1\|\rho^k)-\varepsilon} = \tag{76}$$

$$= \text{Tr}(\Pi^k(\Pi^2)^c(\Pi^1)^c\rho^k(\Pi^1)^c(\Pi^2)^c\Pi^k) +$$

$$+ \text{Tr}(\Pi^k\Pi^2(\Pi^1)^c\rho^k(\Pi^1)^c\Pi^2\Pi^k) + 2^{-nS(\rho^1|\rho^k)-\varepsilon} \leq \tag{77}$$

$$\leq \text{Tr}(\Pi^k(\Pi^2)^c(\Pi^1)^c\rho^k(\Pi^1)^c(\Pi^2)^c\Pi^k) + \text{Tr}(\Pi^2\rho^k\Pi^2) + 2^{-nD(\rho^1\|\rho^k)-\varepsilon} \leq \tag{78}$$

$$\leq \text{Tr}(\Pi^k(\Pi^2)^c(\Pi^1)^c\rho^k(\Pi^1)^c(\Pi^2)^c\Pi^k) + 2^{-nD(\rho^2\|\rho^k)-\varepsilon} + 2^{-nD(\rho^1\|\rho^k)-\varepsilon} \leq \tag{79}$$

$$\leq \dots \leq \text{Tr}(\Pi^k(\Pi^{k-1})^c\dots(\Pi^2)^c(\Pi^1)^c\rho_{\alpha_i}^k(\Pi^1)^c(\Pi^2)^c\dots(\Pi^{k-1})^c\Pi^k) +$$

$$+ \sum_{l=1}^{k-1} 2^{-nD(\rho^l\|\rho^k)-\varepsilon} \tag{80}$$

This means that

$$1 - \delta - \sum_{l=1}^{k-1} 2^{-nD(\rho^l\|\rho^k)-\varepsilon} \tag{81}$$

$$\leq \mathbb{E}[\text{Tr}(\Pi^k(\Pi^{k-1})^c\dots(\Pi^2)^c(\Pi^1)^c\rho_{\alpha_i}^k(\Pi^1)^c(\Pi^2)^c\dots(\Pi^{k-1})^c\Pi^k)] \tag{82}$$

See, that

$$1 - \delta - \bar{a}\min_{1\leq l<k} 2^{-nD(\rho^l\|\rho^k)-\varepsilon} \leq 1 - \delta - \sum_{l=1}^{k-1} 2^{-nD(\rho^l\|\rho^k)-\varepsilon} \tag{83}$$

Because $\varepsilon < \frac{S(\rho_i|\rho_j)}{2}$ this means that if n is enough large then

$$\bar{a} \min_{l \in \{1,2,\dots,k\}} 2^{-nS(\rho^l|\rho^k)-\varepsilon} \le \delta \tag{84}$$

which means that the expectation value of the probability of the good detection is greater than $1 - 2\delta$. This mean the the error (that we detect a wrong mixed state, or all the measure never decide for the first projection) is smaller than 2δ. See that this bound is valid for all possible $1 \le k \le \bar{a}$. Denote by P^k the following operator $\Pi^k(\Pi^{k-1})^c \dots (\Pi^2)^c(\Pi^1)^c$, with this notation at the end of the procedure, our quantum codeword ρ_{α_i} will be in the form $\frac{P^k \rho^o_{\alpha_i} P^{k*}}{\mathrm{Tr}(P^k \rho^o_{\alpha_i} P^{k*})}$

Let see the second step, now we detect the message. Suppose that we detected that our mixed state is ρ^k which mixed state can be generated by $\mathcal{E}^{k_1}, \mathcal{E}^{k_2}, \dots, \mathcal{E}^{k_l}$ $1 \le k_j \le a, l \le a$ and we know that our enemy chosen $\mathcal{E}^o$ so some $k_j = o$. Prepare all $\Pi^{k_j}_{\alpha_i}$ as in Section 3. Now define

$$\check{\pi}_i = span\{ \bigcup_{l:\mathcal{E}^l(\omega)=\mathcal{E}^o(\omega)} \pi^l_i \} \tag{85}$$

where π^l_i is the typical subspace (9) of $\rho^l_{\alpha_i} = \mathcal{E}_l(\omega_{\alpha_i})$ And define the typical projections for the message as in (25), (26)

$$\hat{\pi}_i = span \left\{ P^k|s\rangle - \sum_{j=1}^{i-1} \hat{\Pi}_j P^k|s\rangle, |s\rangle \in \check{\pi}_i \right\} \tag{86}$$

We made our POVM as in Section 3, from these orthogonal projection, with a possible complement with an error labelled subspace. Similarly to Section 3, we define $\hat{\Pi}_{i-} = \sum_{j=1}^{i-1} \hat{\Pi}_j$

For these measurement a similar statement is true as in Lemma 3

$$\mathrm{Tr}(\hat{\Pi}_i P^k \rho^o_{\alpha_i} P^{k*} \hat{\Pi}_i) \ge \left(\mathrm{Tr}(P^k \tilde{\rho}^o_{\alpha_i} P^{k*}) - \mathrm{Tr}(\hat{\Pi}_{i-} P^k \tilde{\rho}^o_{\alpha_i} P^{k*} \hat{\Pi}_{i-}) \right)^2 \tag{87}$$

The proof is exactly the same as Lemma 3. So we can calculate the error probability of the message detection (with the good "mixed" state detection):

$$P_e = \mathbb{E}[\mathrm{Tr}(P^k \rho^o_{\alpha_i} P^{k*}) \frac{1}{M} \sum_{i=1}^{M} 1 - \mathrm{Tr}(\hat{\Pi}_i \frac{P^k \rho^o_{\alpha_i} P^{k*}}{\mathrm{Tr}(P^k \rho^o_{\alpha_i} P^{k*})} \hat{\Pi}_i)] \le \tag{88}$$

$$\le \sum_{i=1}^{M} \frac{\mathbb{E}[1 - \mathrm{Tr}(\hat{\Pi}_i P^k \rho^o_{\alpha_i} P^{k*} \hat{\Pi}_i)]}{M}$$

with the previous statement, and with the Jensen's inequality

$$P_e \le \frac{1}{M} \sum_{i=1}^{M} \mathbb{E}[1 - \left(\mathrm{Tr}(P^k \tilde{\rho}^o_{\alpha_i} P^{k*}) - \mathrm{Tr}(\hat{\Pi}_{i-} P^k \tilde{\rho}^o_{\alpha_i} P^{k*} \hat{\Pi}_{i-}) \right)^2] \le \tag{89}$$

$$\le \frac{1}{M} \sum_{i=1}^{M} \left[1 - \left(\mathrm{Tr}(P^k \tilde{\rho}^o P^{k*}) - \mathbb{E}[\mathrm{Tr}(\hat{\Pi}_{i-} P^k \tilde{\rho}^o_{\alpha_i} P^{k*} \hat{\Pi}_{i-})] \right)^2 \right] \tag{90}$$

Because $P^k\rho^o P^{k^*} \geq 1-2\delta$ and $\mathrm{Tr}(\rho^o - \tilde{\rho}^o) \geq 1-\delta$ the first term is greater than $1-3\delta$ as in proof of Theorem 1. The second term is

$$\mathbb{E}(\mathrm{Tr}(\hat{\Pi}_{i-}P^k\tilde{\rho}^o_{\alpha_i}P^{k^*}\hat{\Pi}_{i-})) = \mathbb{E}(\mathrm{Tr}(\hat{\Pi}_{i-}P^k\tilde{\rho}^o_{\alpha_i}P^{k^*})) =$$

(91)

$$\mathbb{E}(\mathrm{Tr}(\sum_{j=1}^{dim(\hat{\Pi}_{i-})}|b_j\rangle\langle b_j|P^k\tilde{\rho}^o_{\alpha_i}P^{k^*})) \leq \mathbb{E}(\sum_{j=1}^{dim(\hat{\Pi}_{i-})}\langle b_j|\Pi^k\tilde{\rho}^o_{\alpha_i}\Pi^k|b_j\rangle) \leq dim(\hat{\Pi}_{i-})\|\Pi^k\tilde{\rho}^k\Pi^k\|$$

(92)

because $P^k\tilde{\rho}^o_{\alpha_i}P^{k^*} \leq \Pi^k\tilde{\rho}^o_{\alpha_i}\Pi^k$, and we assumed that $\mathcal{E}_o(\omega) = \mathcal{E}_k(\omega)$. Now we have to evaluate $dim(\hat{\Pi}_{i-})$

$$dim(\hat{\Pi}_{i-}) \leq \sum_{j=1}^{l}\sum_{s=i+1}^{M} dim(\Pi^{k_j}_{\alpha_s}) \leq aM2^{nS(\rho|\alpha)} = a2^{n(R+S(\rho|\alpha))}$$

(93)

We use (23) and get

$$\mathbb{E}(\mathrm{Tr}(\hat{\Pi}_{i-}\Pi^k\tilde{\rho}^o_{\alpha_i}\Pi^k\hat{\Pi}_{i-})) \leq a2^{-n[S(\rho^o)-S(\rho^o|x)-R-\varepsilon]}$$

(94)

which is smaller than δ if n is large enough, so

$$P_e \leq \frac{1}{M}\sum_{i=1}^{M}[1-(1-3\delta-\delta)^2] < 8\delta \leq 8a\varepsilon$$

(95)

Because $\varepsilon < \frac{\gamma}{8a}$ with this the theorem is proved. ∎

5 II. Use: Practical considerations

One could think that, after the articles of Schumacher or Holevo [3], [4] that we can communicate classical data through a quantum channel optimally. However this is true only in theory, because to measure a POVM with many output (the needed output of the POVM grows exponentially in n the length of the codeword) is very difficult in practice. But as we will see, this is not the case in the von Neumann measurement, we will give a detection algorithm - a sequence of measure - where the number of outcomes of the measures are always 2.

We introduce the following notation for $1 \leq i < j \leq M+1$

$$D_{\{i,j\}} = \tilde{\Pi}_i + \tilde{\Pi}_{i+1} + \cdots + \tilde{\Pi}_j$$

(96)

For simplicity, suppose that $M+1 = 2^k$. Then the detection algorithm can be the following: First we measure a von Neumann measurement states from $D_{\{1,(M+1)/2\}}, D_{\{(M+1)/2+1,M+1\}}$. In every measurement, if the result is the first operator, we give 0, if the second we give 1.

Now we measure again. Of course on a quantum state that is modified by the previous measurement. In each next step we half the interval of the previous measurement. If our measurement gave the $\{i, j\}$ our measurement will states from $D_{\{i, i+(j-1)/2\}}, D_{\{i+(j-1)/2+1, j\}}$. For example the second step looks like follows: If we measured 0 then our measurement will states from $D_{\{0, ..., (M+1)/4\}}, D_{\{(M+1)/4+1, ..., (M+1)/2\}}$, if the previous measurement had gave the result 1 then our measurement will states from :
$D_{\{(M+1)/2+1, ..., 3(M+1)/4\}}, D_{\{3(M+1)/4+1, ..., (M+1)\}}$. At the end the 0-s and 1-s give the number of the message in binary form. If we get only 1-s then we declare error.

See that the probability of the good detection not changes. Suppose that we send the first message, then the first measure will give the good result with probability

$$\text{Tr}(D_{\{\{1, (M+1)/2\}}\rho_{\alpha_1} D_{\{1, (M+1)/2\}}) \tag{97}$$

and will the state will change to

$$\frac{D_{\{1, (M+1)/2\}}\rho_{\alpha_1} D_{\{1, (M+1)/2\}}}{\text{Tr}(D_{\{1, (M+1)/2\}}\rho_{\alpha_1} D_{\{1, (M+1)/2\}})} \tag{98}$$

The second measurement will be good with probability

$$\text{Tr}\left(D_{\{1, (M+1)/4\}} \frac{D_{\{1, (M+1)/2\}}\rho_{\alpha_1} D_{\{1, (M+1)/2\}}}{\text{Tr}(D_{\{1, (M+1)/2\}}\rho_{\alpha_1} D_{\{1, (M+1)/2\}})} D_{\{1, (M+1)/4\}}\right) \tag{99}$$

But $D_{\{1, (M+1)/4\}} < D_{\{1, (M+1)/2\}}$ so this simplify to

$$\text{Tr}\left(\frac{D_{\{1, (M+1)/4\}}\rho_{\alpha_1} D_{\{1, (M+1)/4\}}}{\text{Tr}(D_{\{1, (M+1)/2\}}\rho_{\alpha_1} D_{\{1, (M+1)/2\}})}\right) \tag{100}$$

And the state change to

$$\frac{D_{\{1, (M+1)/4\}}\rho_{\alpha_1} D_{\{1, (M+1)/4\}}}{\text{Tr}(D_{\{1, (M+1)/4\}}\rho_{\alpha_1} D_{\{1, (M+1)/4\}})} \tag{101}$$

So the probability that the first two measurement was true is

$$\text{Tr}(D_{\{\{1, (M+1)/2\}}\rho_{\alpha_1} D_{\{1, (M+1)/2\}}) \text{Tr}\left(\frac{D_{\{1, (M+1)/4\}}\rho_{\alpha_1} D_{\{1, (M+1)/4\}}}{\text{Tr}(D_{\{1, (M+1)/2\}}\rho_{\alpha_1} D_{\{1, (M+1)/2\}})}\right) = \tag{102}$$

$$\text{Tr}(D_{\{1, (M+1)/4\}}\rho_{\alpha_1} D_{\{1, (M+1)/4\}}) \tag{103}$$

with keep going this train of thought we can see, that at the end that the probability that all of the measurement was good is not else, than

$$\text{Tr}(D_{i,i}\rho_{\alpha_i} D_{i,i}) = \text{Tr}(\tilde{\Pi}_i \rho_{\alpha_i} \tilde{\Pi}_i) \tag{104}$$

which the same as the error probability in Section 3. Of course this procedure can be generalised to case when we have finite possible outcomes.

This means that it is possible to classically code/decode classical information with quantum apparatus in an optimal way in linear time. This is a quite surprising result, because in classical information theory to reach the Shannon limit in polynomial time is an unresolved problem (the best result needs $n^{\log(n)}$ time). Usually the classical information theory is considered as a part of quantum information theory, which would mean that optimal decoding of classical channel in linear time is possible. This means that if a quantum machine can perform arbitrarily von Neumann measurement with only two possible outcome, then this machine can solve non-polynomial classical problems in linear time.

Now we will show how can be a classical message through a quantum apparatus decoded. Suppose that there is a classical setup with a discrete memoryless channel. There is a given state transition matrix $W(y|x)$ (with input output alphabet $1, 2, \ldots, l$ $1, 2, \ldots, d$) and a given optimal input distribution P. Now model the classical system with a quantum one. Let define for each x $\rho_x = diag(W(\cdot|x))$ (where $diag(W(\cdot|x))$ denotes a diagonal matrix we get from the output distribution provided by x in another form $\sum_{a \in \{1, \ldots, d\}} W(a|x) E_{a,a}$). In these case all the classical and all the quantum information quantities are equivalent ($\chi(\mathcal{E}) = C(W)$, $H(\cdot) = S(\cdot)$). Then we know from Section 3 that there exist 2^{Rn} piece of sequence that with ρ_{α_i} quantum codewords we can optimally communicate. Compute the optimal von Neumann measurement as in Section 3. Now we use the α_i sequence as an input codeword for our classical channel, and decode the classical channel as follows: We get the classical signal, we coded into quantum sequences, we perform the measurement, after that we get the number of the message was sent, so we decoded the message (in linear time as in the beginning of these section).

Denote the output signal of α_j by $\underline{y}(j)$. Denote the i-th component of $\underline{y}(j)$ by $y_i(j)$ We get the signal $\underline{y}(j)$ and code every symbol of it, into a quantum in the following way:

$$y_i(j) \rightarrow E_{y_i(j),y_i(j)} \tag{105}$$

which means if we get the first symbol of my output alphabet we code into a quantum represented by $E_{1,1}$, where $E_{i,j}$ denotes the matrix with 1 in the i-th row j-th column and 0 elsewhere. Denote these quantum sequence by $\mu_{\underline{y}(j)}$ Now see that the error probability of the event that the i-th message was wrongly decoded:

$$\mathbb{E}[1 - \text{Tr}(\mu_{\underline{y}(j)})] \tag{106}$$

We have to take the expectation value because the output sequence $\underline{y}(j)$ can varied. It can be easily proved that $\mathbb{E}[\mu_{\underline{y}(j)}] = \rho_{\alpha_j}$. So the average error probability is same as in Section 3. Which means that classical messages can be decoded by quantum apparatus in linear (nR) time.

Proof:of $\mathbb{E}[\mu_{\underline{y}(j)}] = \rho_{\alpha_j}$ proof with total induction on n for $n = 1$ the

16

statement is true by the definition. Suppose it is true for $n-1$ the for n

$$\mathbb{E}[\mu_{\underline{y}(j)}] = \sum_{\underline{y}\in\{1,\ldots,d\}^n} \prod_{i=1}^n W(y_i|\alpha_j(i))\mu_{\underline{y}} \tag{107}$$

$$= \sum_{\underline{y}\in\{1,\ldots,d\}^{n-1}} \prod_{i=1}^{n-1} W(y_i|\alpha_j(i))\mu_{\underline{y}_1^{n-1}} \otimes \sum_{a\in\{1,\ldots,d\}} W(a|\alpha_j(n))E_{a,a} \tag{108}$$

where $\underline{y}_1^{n-1}$ denote the first $n-1$ symbol of $\underline{y}$. But by the definition the last quantity $\sum_{a\in\{1,\ldots,d\}} W(a|\alpha_j(n))E_{a,a}$ is not else than $\rho_{\alpha_j(n)}$ so

$$\mathbb{E}[\mu_{\underline{y}(j)}] = \sum_{\underline{y}\in\{1,\ldots,d\}^{n-1}} \prod_{i=1}^{n-1} W(y_i|\alpha_j(i))\mu_{\underline{y}_1^{n-1}} \otimes \rho_{\alpha_j(n)} \tag{109}$$

but for $n-1$ the statement is true, so

$$\mathbb{E}[\mu_{\underline{y}(j)}] = \rho_{\alpha_j} \tag{110}$$

∎

A   proof of Lemma 2

The proof based on typical sequences. These definition is a simplified/modified version of [6].

Definition 1 (Typical sequence) *For a given probability distribution P on $\{1,2,\ldots,d\}$ an $\underline{x}\in\{1,2,\ldots,d\}^n$ sequence is called P-typical with constant δ, if*

$$\left|\frac{1}{n}N(a|\underline{x}) - P(a)\right| \leq \frac{\delta}{\sqrt[4]{n}} \quad \text{for every} \quad a\in\{1,2,\ldots,d\} \tag{111}$$

where $N(a|\underline{x})$ means the number occurrences of a in sequence $\underline{x}$ and, in addition no $a\in\{1,2,\ldots,d\}$ with $P(a)=0$ occurs. The set of such sequences will be denoted by $T_{[P]\delta}^n$ or simply $T_{[P]}$.

Remark 3 *If a sequence $\underline{x}$ is P-typical as above, then*

$$\left|-\sum_{a=1}^d \frac{N(a|\underline{x})}{n}\log(P(a)) + \sum_{a=1}^d P(a)\log(P(a))\right| \leq \frac{Kd\delta}{\sqrt[4]{n}} \tag{112}$$

if δ is small enough, because if $P(b)$ is 0 then $N(b|\underline{x})=0$ so $\frac{N(b|\underline{x})}{n} - P(b) = 0$ so the b-th element of the sum will be 0. If $P(a)>0$ then $\log(P(a))$ is finite, so $max_{a:P(a)>0}[-log(P(a))] = K$ is finite, so the above sum is smaller than $\frac{Kd\delta}{\sqrt[4]{n}}$.

17

Lemma 4 *For every distribution P on $\{1, 2, \ldots, d\}$, and for every $\beta > 0$*

$$P^n(T^n_{[P]_\delta}) \geq 1 - \beta \tag{113}$$

if n is large enough.

Proof: If $\underline{X} = X_1, X_2, \ldots, X_n$ is an i.i.d. random sequence with distribution P then the random variable $N(a|\underline{X})$ has the expectation value $nP(a)$ and variance $nP(a)(1 - P(a)) \leq \frac{n}{4}$. Thus by the Chebishev's inequality

$$\Pr\left\{ \left| N(a|\underline{X}) - n(Pa) \right| > n\frac{\delta}{\sqrt[4]{n}} \right| \leq \frac{1}{4\sqrt{n}\delta^2} \right\} \tag{114}$$

for every $a \in \{1, 2, \ldots, d\}$. From this the assertion follow. ∎

of Lemma 2: with these typical sequences we can make typical subspace as follows:

Let $\sum_{i=1}^{d} \lambda_i |u_i\rangle\langle u_i| = \rho$ be a spectral decomposition of ρ. Now define P as $P(a) = \lambda_a a \in \{1, 2, \ldots, d\}$ (In this case $H(P) = S(\rho)$, and let n be large enough to verify 113 with $\beta = \varepsilon/2$. Now we define Π the typical projection of $\rho^{\otimes n}$ by as follows

$$\Pi = \sum_{\underline{x} \in T^n_{[P]_\delta}} |u_{x_1}\rangle\langle u_{x_1}| \otimes |u_{x_2}\rangle\langle u_{x_2}| \otimes \cdots \otimes |u_{x_n}\rangle\langle u_{x_n}| \tag{115}$$

See that if a sequence $\underline{x}_1$ differs from $\underline{x}_2$ then the minimal projection generated by $\underline{x}_1$ is orthogonal to the minimal projection generated by $\underline{x}_2$.

For this projection $\hat{\Pi}$ the assertions (22) and (23) of Lemma 2 are valid. See the first assertion

$$\mathrm{Tr}(\hat{\Pi}\rho^{\otimes n}\hat{\Pi}) = \mathrm{Tr}(\hat{\Pi}\rho^{\otimes n}) = \tag{116}$$

$$= \mathrm{Tr}(\sum_{\underline{x} \in T^n_{[P]_\delta}} |u_{x_1}\rangle\langle u_{x_1}|\rho \otimes |u_{x_2}\rangle\langle u_{x_2}|\rho \otimes \cdots \otimes |u_{x_n}\rangle\langle u_{x_n}|\rho) =$$

$$\tag{117}$$

$$= \sum_{\underline{x} \in T^n_{[P]_\delta}} \prod_{i=1}^{n} \langle u_{x_i}|\rho|u_{x_i}\rangle = \sum_{\underline{x} \in T^n_{[P]_\delta}} \prod_{i=1}^{n} \lambda_{x_i} = \tag{118}$$

$$= \sum_{\underline{x} \in T^n_{[P]_\delta}} \prod_{i=1}^{n} P(x_i) = P^n(T^n_{[P]_\delta}) \geq 1 - \beta \tag{119}$$

if n is large enough $n > N_1$, and the last row comes from the definition of P and the previous Lemma.

Observe that (23) is true, because of Remark 3. See that spectrum of $\hat{\Pi}\rho\hat{\Pi}$ is equal with

$$spect(\hat{\Pi}\rho\hat{\Pi}) = \{ \prod_{a=1}^{d} \lambda_a^{N(a|\underline{x})}, \underline{x} \in T^n_{[P]_\delta} \} \tag{120}$$

where

$$\prod_{a=1}^{d} \lambda_a^{N(a|\underline{x})} = 2^{-n \sum_{a=1}^{d} -\frac{N(a|\underline{x})}{n} \log(\lambda_a)} \tag{121}$$

and from Remark 3 we know that for all $\underline{x} \in T^n_{[P]_\delta}$

$$\prod_{a=1}^{d} \lambda_a^{N(a|\underline{x})} \leq 2^{-n(S(\rho) - \frac{Kd\delta}{\sqrt[4]{n}})} \tag{122}$$

where $\frac{Kd\delta}{\sqrt[4]{n}}$ is smaller than β if n large enough.

We know from [1] that, if n is large enough, there is an another projection $\tilde{\Pi}$ which satisfy (22), (24). Now the projection which satisfy all the assertion of the Lemma is given by $\Pi = \hat{\Pi} \wedge \tilde{\Pi}$, because

$$\mathrm{Tr}(\Pi^c \rho \Pi^c) \leq \mathrm{Tr}(\hat{\Pi}^c \rho \hat{\Pi}^c) + \mathrm{Tr}(\tilde{\Pi}^c \rho \tilde{\Pi}^c) = 2\beta \tag{123}$$

$$\|\Pi \rho \Pi\| \leq \|\hat{\Pi} \rho \hat{\Pi}\| \leq 2^{-n(S(\rho) - \beta)} \tag{124}$$

$$\mathrm{Tr}(\Pi \sigma^{\otimes n} \Pi) \leq \mathrm{Tr}(\tilde{\Pi} \sigma^{\otimes n} \tilde{\Pi}) \leq 2^{-n(D(\rho \| \sigma) - \beta)} \tag{125}$$

■

References

[1] Fumio Hiai and Dénes Petz "The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability" in Communications in Mathematical Physics vol. 143 pp. 99-114 (1991)

[2] B. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, W. K. Wooters "Classical information capacity of a quantum channel" in Physical Review A Vol. 54. pp 1869-1876.(1996)

[3] B. Schumacher, "Sending classical information via noisy quantum channels" in Physical Review A Vol. 56. No.1 pp 56-138.(1997)

[4] A. S. Kholevo, "Some estimates of the information transmitted by quantum communication channel" in Probl. Peredachi Inform. vol. 9, 3-11 (1973)(rus)

[5] A. S. Holevo "Capacity of the quantum channel with general signal states" in IEEE Transaction on Information Theory Vol. 44. iss. 1 pp. 269-273 (1998)

[6] I. Csiszár J. Körner, *Information Theory Coding Theorems for Dicrete Memoryless Systems* Akadémiai kiadó 1986.